Spatial Informatics
Prof. Soumya K. Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 05
Spatial Data Models (4)

Hi. So, we will continue our discussion on Spatial Data Models, if you recollect couple of last lectures. So, what we are discussing about that how a spatial information can be modeled right. So, our objective is to given a spatial data or spatial scenario for our cases again I am repeating it is a geospatial data, how do I model it right. As we as a example we are trying to see that if I have a data set consisting of say road, river different type of land use land cover and say build up areas how do I model it right. So, our basic objective all those things is that to query on this data all right.

I need to find out the shortest path between location A and location B like that is may be one very trivial or but most one of the popular operations on the spatial data, then while considering thing I need to know that what are the I need if it is the shortest path then I have to find out the path between location A B considering different category of roads; some may be state highways, some may be national highways, some may be street road and so and so forth.

So, these roads need to be mapped right unless these roads are available at a platform then your search engine will not work right. Suppose I am using some application which can basically starts this type of thing then I need to have some this application data in at some level right.

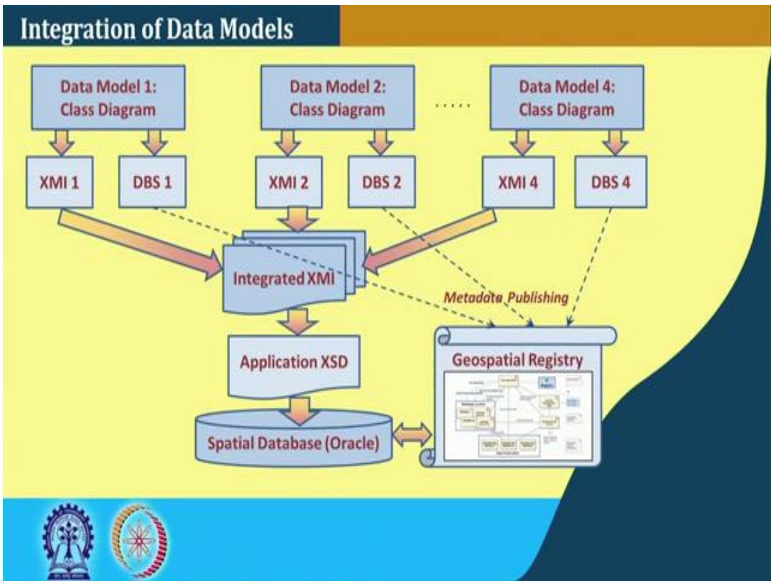
So, that means, in other sense from the source of the data needs to be modeled then they talk to each other right and the things become more complicated if my whole say transport if we even if you consider the transportation domain and my the transportation network consists of say road, rail, air and so on and so forth. So, that things becomes more complicated than finding a point A to point B becomes a more complicated issue because these are or with different repositories, a different state repositories with different administrative control and unless they kept in a appropriately modeled into appropriate thing then it is difficult to query on the things.

So, our objective of this spatial modeling is that, how I can achieve a some sort of a standardization which helps me to talk to each other, nevertheless we need to fall back the whole thing into a database.

Unless there is a database in place then querying, retrieving, storing all those becomes very difficult right that we everybody understands that without database with large set of large scale data handling will be much difficult right. So, we will later on we will subsequently we will discuss about something about spatial database and then that how this spatial query and things are there and how they different.

But so, this slide we have seen earlier right in our earlier lectures and talks.

(Refer Slide Time: 03:39)



So, what we are having that different say repositories right. So, there are different data models suppose this is road, this is say if you consider the transportation itself, road, rail that maybe waterways or something etcetera and so and so forth there are different type of things right. So, now, these different repositories I need to model right. I need to model that how my data is there. So, or I want to if I want to share then I have to model this thing right. So, how who can model the base? The person who is maintaining this the domain knowledge. So, one of the standard modeling what we have seen is the UML right.

So, if I can have a UML model, then going to the next step is creating a some sort of a XMI sort of things right as if you recollect or remember; so, XMI type. From the XMI we can create from the XMI I can have this integration of the different XMIs right different sources XMIs are there and then I can have a application schema or application XSD right. So, this is these are all XML based technologies I believe that most of you are conversant with or in the subsequent lectures, we will discussed about also spatial web services where we talk about XML and other services how things will be there.

But the basis is that schema definition or I have a structural definition of the things right. The beauty of the thing is that I am not sharing the data, the data remains still in place right I only trying to share a their structure or in other sense I want to create a schema which is basically for those applications I am looking for right. Like if my application is related to finding different sort of shortest path and etcetera is the thing then I want a schema which my query on those things can work on it. Now I want to ensure that who are the supplier of the data, how do I ensure from the thing? From who? The stakeholders or the data repository, I want to what we are looking for that whether we can share their structure right.

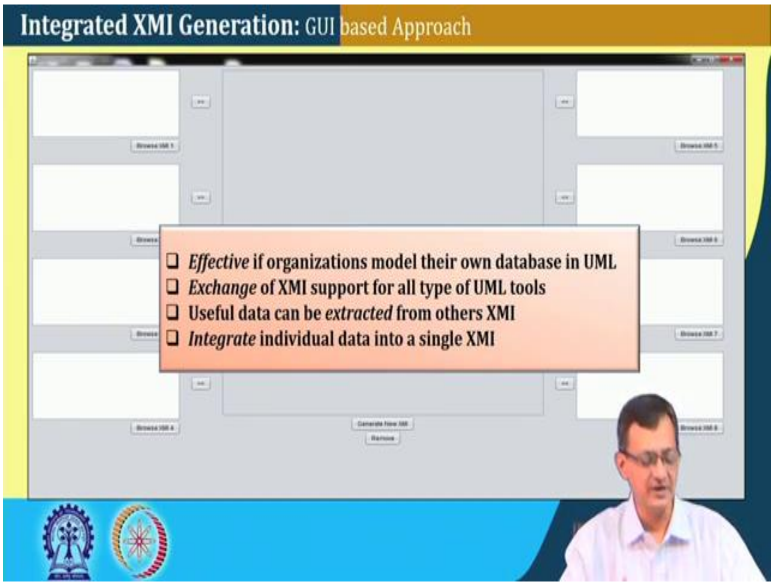
Suppose I am looking for a road network or rail network, if the rail network data is not there then my query will fail or the road network data. So, I need to find out that if I want to find that Kharagpur up to street road, street level roads then I need to have those data in place. So, first of all a prioriand I do have this schema in place now if it is there then I can put it to a spatial database right. So, oracle is one oracle spatial these days we have mysql, mysql spatial extension postfix is there and we there are different type of proprietary tools or more which are there are several supports are there right.

Now a application or user or a application what it want to know? It wants to know this sort of a registry right first where I will consult, I will consult to the registry that where are the data sets are there if b is the data sets are there, then in find out the find out the application schema on the things right.

So, things are clear like I have different repositories which are the data stored in different things, they create their model which they if you want to change in say in UML, then want to have a XMI sort of thing from XMI different XMI integration we create a XSD and this XSD helps me in creating the underlining database structure or DDL type of things right data definition or that database structure of the underlying spatial database. If this is done then I can basically query on the things right. Now whenever the f for whenever a repository say highway authority or say transport authority which road transport authority, whenever they are coming up with a new road or new type category of road they populate on the on their registry right which is kept in a centralized registry.

So, as a application user, I consult that registry whether this is things are available then I query on the things right so, but see unless I query the data has not transferred. So, data still remain in space that is in place right whoever is maintaining and type of things that is important otherwise maintainability of the data is will be extremely difficult.

(Refer Slide Time: 08:41)

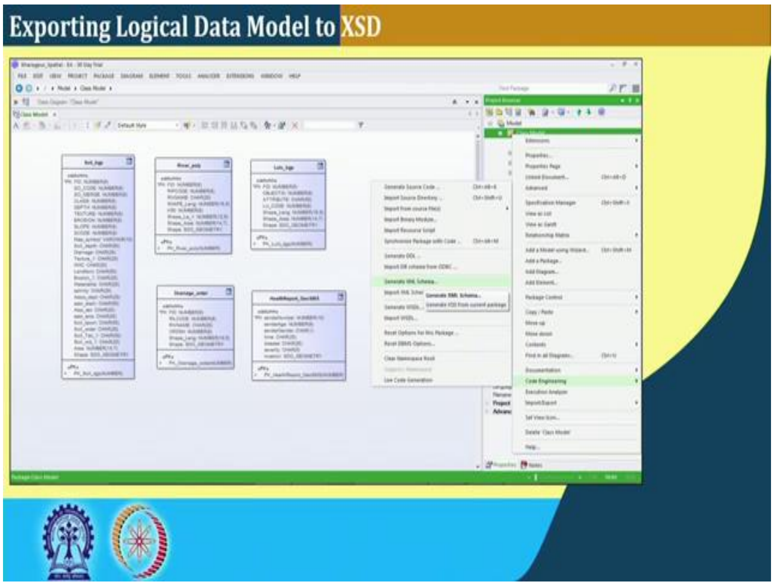


Now, this also we have seen that I can have some GUI based that finding that who what are the different data sets.

(Refer Slide Time: 08:50)

Like in this case that a typical BNK metal road non metal road and type of things right.

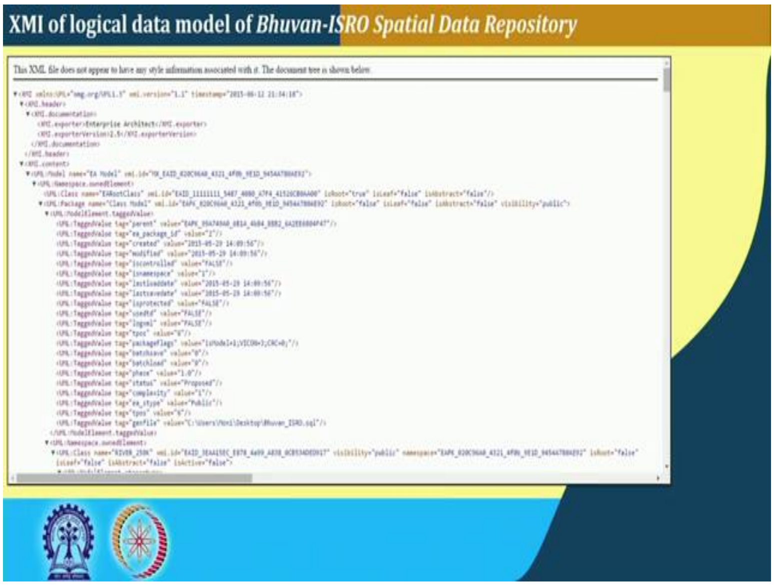
(Refer Slide Time: 08:55)



And then we have a logical data model UML to the XMI right. So, the a say there are different type of classes right.

This is a these are some example classes taken from some data sets, but there can be based on your particular domain. So, here it is say soil kgp or soil map a river map, drainage, LULC that Land Use and Land Cover data health report data that means, what are the different health related issues etcetera. So, if it is a health related things there can be different type of datasets based on the things which are being collected or which you need to you need to query about. Now this is the I have this particular UML right

(Refer Slide Time: 09:50)

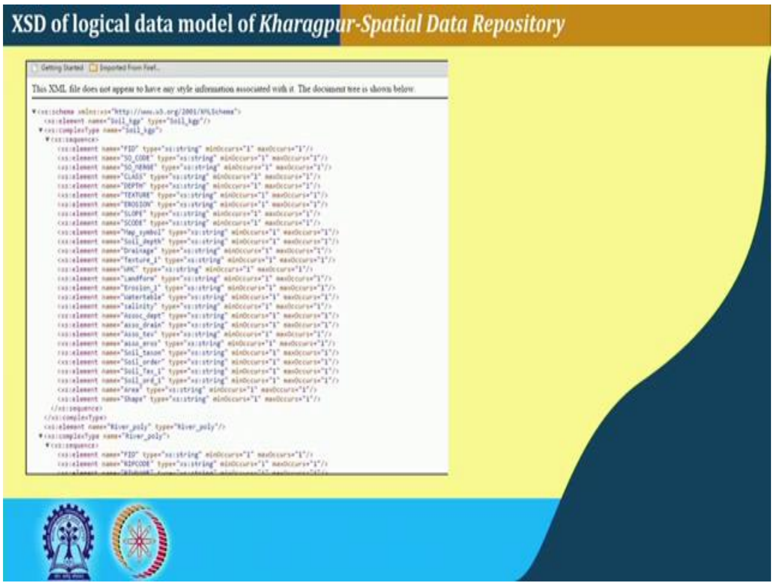


From that if it is there like in this Kharagpur spatial data repository we can create that XMI right which is a XML based technology and interchangeable things right. So, these XMI of logical data model of the Kharagpur, right. Once this is there now this is machine readable right this is a machine readable, then I can parse this data in the in my systems and extract knowledge out of it right.

So, there are there are different way of representation as of now we please we will not go or we will not break our head on that what are the structure etcetera our. We need to know what is the basic philosophy of integration of data set from different repositories right. So, those who are interested we can go you can go deep into the things these are all standard technologies which are being placed in different other things but the for spatial data these are being press recently right.

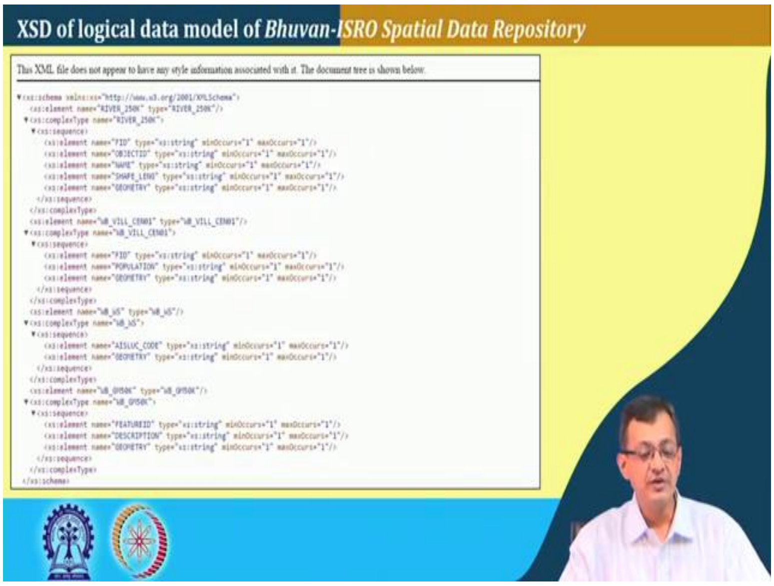
So, similarly another data set we have taken from Bhuvan ISRO spatial data repository right. So, one is that Kharagpur another is that another repository there can be different type of repository where we can have right and then we can model to the this logical model. So, which we have this different type of data set and created this class diagram we can export as a XSD. So, this combined model can go as a XSD right.

(Refer Slide Time: 11:24)



So, XSD of the logical data model of Kharagpur spatial data repository right. Those who are accustomed with x XSD and type of things can appreciate this those who are not so, a custom will be having one couple of classes where we will quickly though quickly browse what will give you the some of the basic working of the things anyway XSD of this particular Kharagpur spatial data we can have that XSD of the other repository in this case Bhuvan ISRO spatial repository.

(Refer Slide Time: 11:53)



See these are the two repositories what we are taking are maintained by two different organization or agencies for their own purpose. What they are sharing it is the structure, by looking at the structure I can create a unified or schema or what we say application schema which is needed for that queries. Once that is there I can basically query on those and the data can be extracted from the repositories in a in a service mode right spatial service mode. How that we will see in the subsequent lectures, but if I can extract these data in a subsequent service mode then I can combine them and execute right.

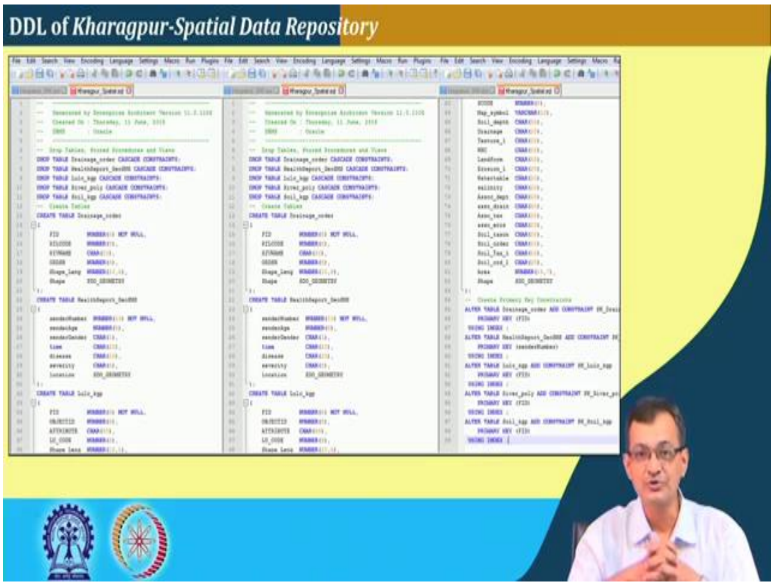
The example we try to fall back in this case is like you are booking rail ticket an air ticket. So, what you are having a something a broker E broker right like IRCTC or something. So, that it that it goes to this railway database and extract those information the particular train etcetera, then also if you want to pay by different option like credit, debit, net banking etcetera it interface with that right and then compositely it resolve the challenge right resolve the query so, that the ticket is booked or something right.

Now this is our service driven. Now for that what we require? I require the schema right for which data is there if it is a flight related data then these different companies will come into play different airlines will come into play there are different timetables different pay structure and type of things. So, that need to be integrated.

So, what they share is the structure right. So, different this E at operators or E tourists operator or E ticket operators they do not have plane or train or something right, but they have those repository and make you available. So, what they maintain is basically a registry over which you can fare queries like I want to go from here to here, what is the minimal path if it is a multi hub then what are the etcetera right. I can fly then book a car book a hotel and all those things can be combined right.

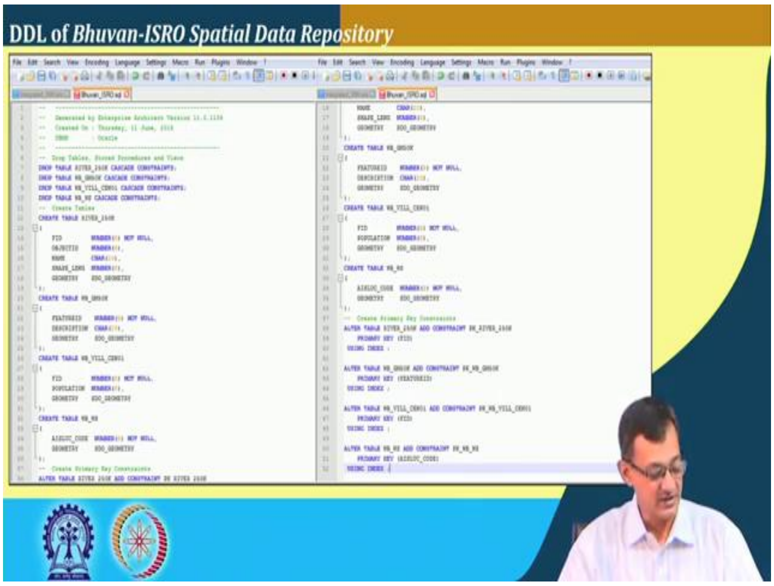
Here if I have spatial data like this my query expand with different requiring different data structure I need something like this right. So, that is the XSD of the Bhuvan ISRO spatial data repository right so; that means, I have two repositories and then I can has export the particular logical structure.

(Refer Slide Time: 14:48)



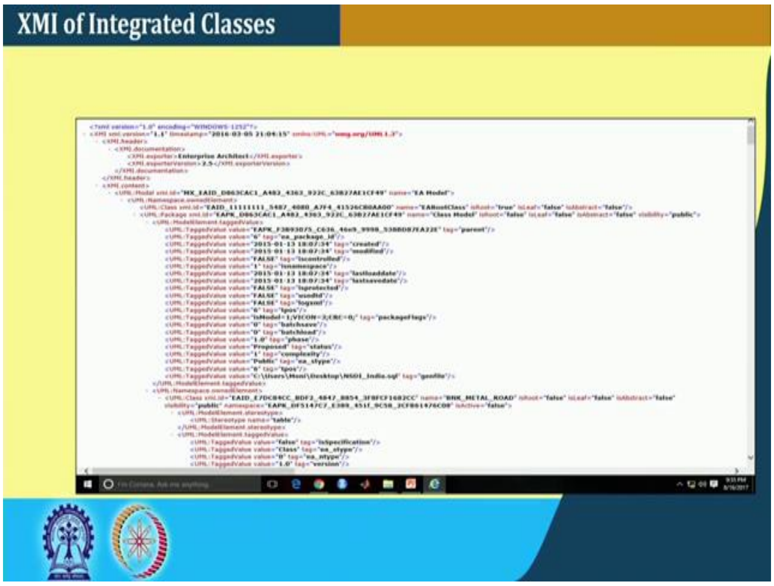
And then I can have a DDL of the Kharagpur spatial depository right. Now we are falling back to this database schema right. So, these are the different database schema for different type of this particular Kharagpur repositories right. So, whatever the that repository is having.

(Refer Slide Time: 15:11)



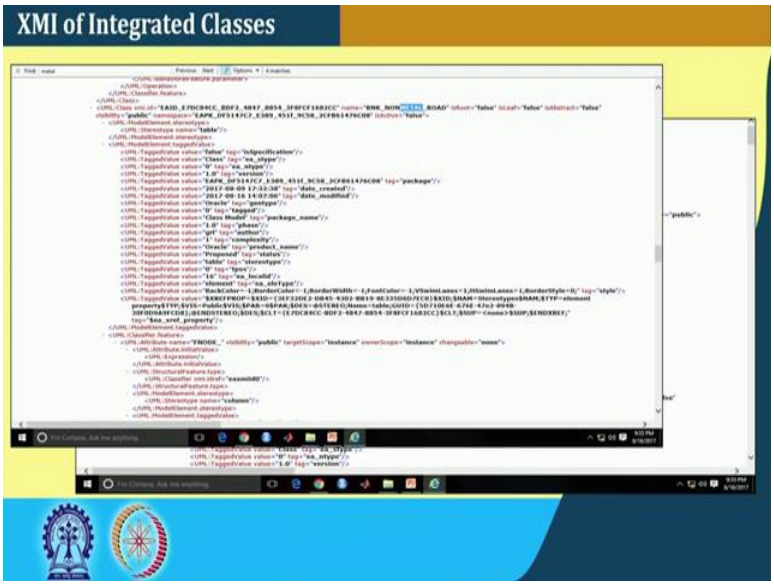
Similarly, we have that DDL of the Bhuvan ISRO spatial repositories right. So, that it is now end of the database schema and this is that XMI of the integrated class right.

(Refer Slide Time: 15:21)



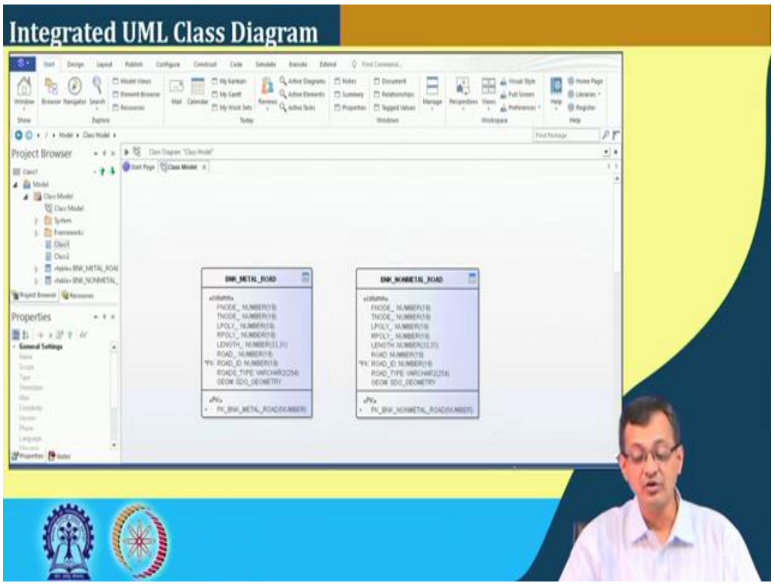
So, we have this XMI of the integrated class of having both Kharagpur and a Kharagpur repository and this Bhuvan repository right. So, again I am not we are not going deep into the things that what are the structure means what etcetera, but we can able to generate a integrated XMI right.

(Refer Slide Time: 15:50)



So, integrated classes etcetera of different repositories we can generate.

(Refer Slide Time: 15:59)

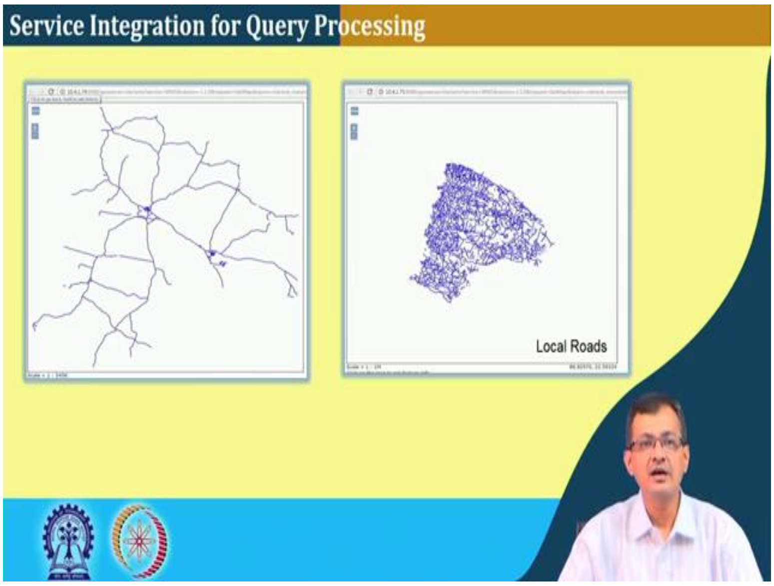


And this helped me in to come up with a some sort of a integrated UML model or we can have a UML model. Like here we have that metal road and nonmetal road from the two different sources and we have a integrated model of these things right. So, this is important.

So, that I can have now a model which inherits from two different repositories model or indifferent repositories model; now if it is there then we can have different queries into the place right. So, again before showing some snapshots and type of things which will again later on. So, in different in a in some sort of a demo type of things.

So, our again I am just repeating because this philosophy is important that I is in individual repositories are modeled by UML model create XMI then combine them and create a XSD and I can have a now a integrated application schema or integrated XSD which helps me in creating the database right. I can create individual database parse say also right I can have a integrated UML model which allows me to create this database.

(Refer Slide Time: 17:27)



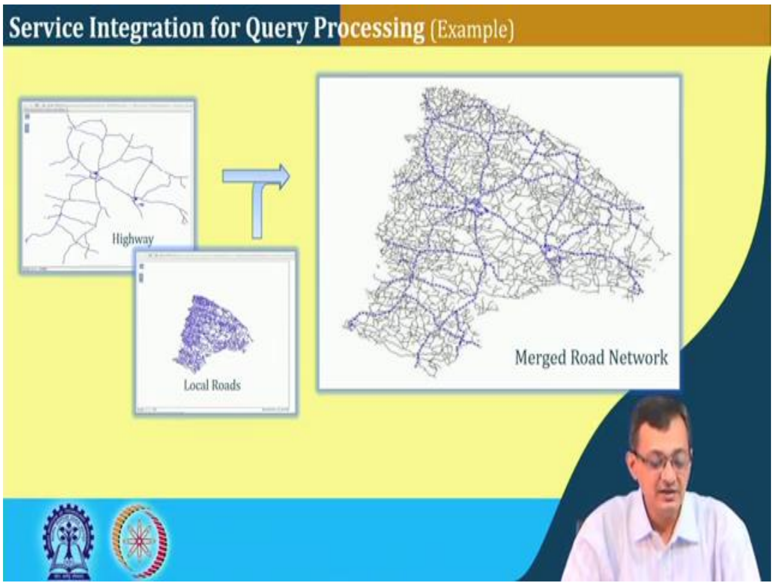
Now, see let us look at something right again two type of network this may be this level is missing. So, these are the highway right say this is a highway region is a Bankura district of the thing.

So, and these are all local roads. So, these are state highways and national highways whichever passing and these are the local roads of that particular region. Now these things are maybe in the two things local roads are maintained by maybe the district transport authority whereas, the highways and state highways are maintained by say state highway authority or national highway national authorities.

Now if I want to find something that if I want to go from say this place to some this place either there is road is not there if I look into things, but if you look into the this particular local roads so, there is a path of going there right. So, finding a path if I am driving down the things and go to particular a village or village headquarter, then I need to find out that where are the paths are there.

So, this need to be combined right. Now one way is that I get the data combined at locally and then query fine next time again you require a fine you get it the data again because you do not know that whether the paths etcetera are changed or something is blocked etcetera. So, one way is that keep the data at place what you require at that for the integration is that integration of the structure and then wearing where firing the query you extract the data from the data sources right.

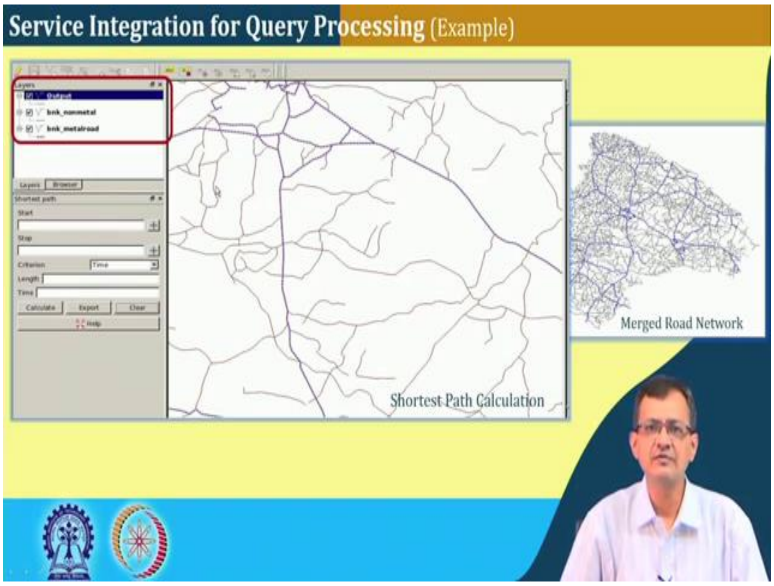
(Refer Slide Time: 19:07)



Now, this is the thing like that was the highway and local road and we have a view of the merged network right. We at the top title we are mentioning service integration of the query processing. So, it is service level integration right. So, using spatial services we are extracting the data and integrating in at the service level right.

So, that is important that service level integration how that will look into the things, those who are having knowledge or on the spatial web services and etcetera and not only that normal web services and. So, service oriented architecture fine otherwise we will have a we will we will see that how this spatial web service works in subsequent lectures. See this is the beauty of the things. So, to here we are displaying, but we can query also. So, we have highway local roads and then I merge them or into a things for the display purpose. So, merge road network right.

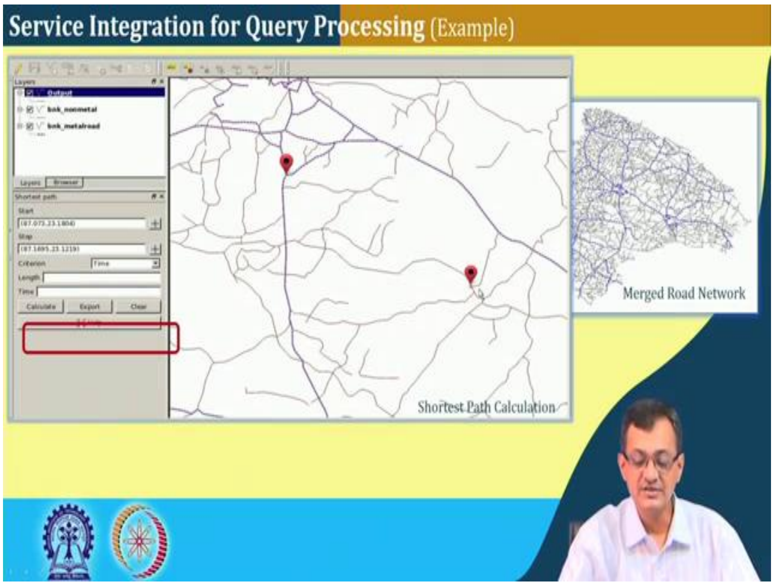
(Refer Slide Time: 20:09)



So, one very traditional application is finding a shortest path right finding the shortest path between location A and location B and it may have it may have both these local roads and state roads or state highways national highways etcetera.

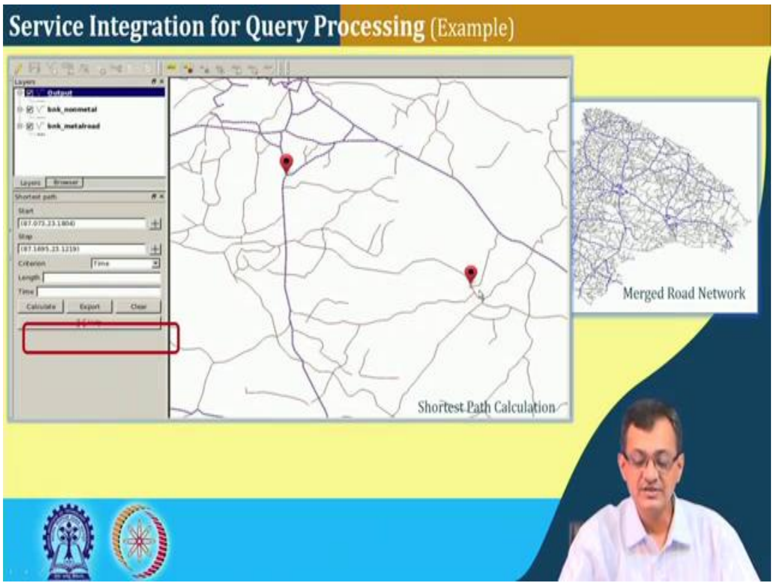
So, somewhere integration needs to be there right. So, finding that particular path between shortest path between location, A and B right. So, this sort of service integration for the process queries on the things.

(Refer Slide Time: 20:46)



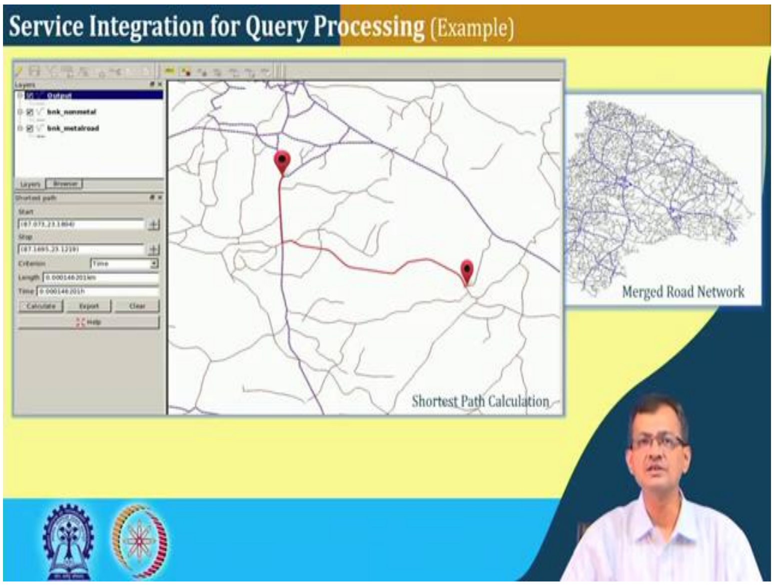
So, say this is my location A or source for that particular things.

(Refer Slide Time: 20:55)



And then I have another location here at destination and I want to find out the shortest path right. Now here in this particular case these are all polylines right polylines and there are they can be considered as a graph in CS point of view. Now if it is a graph, then I have that different graph label algorithms to work on those things right. So, then I find out a shortest path between these two things two particular locations like I have this particular path for that matter right.

(Refer Slide Time: 21:31)



So, it is it may be the shortest path or we can have different other type of things.

Now, this is a this application source that my basically my application is finding the shortest path between two location A and B, for that I require road network from two sources it could have been so, that I have multimodal transport mechanisms. So, I can I am ready to use multimodal; that means, I have road network, rail network maybe water bodies right. So, there are different sources of this connectivity right or and then in order to find out the shortest path, I need to have integrate them into a common base right.

So, that is the that where that things come into play. So, that what we want that that structurally or the schema wise integration and the result is schema is can give me shortest path or that type of operations right.

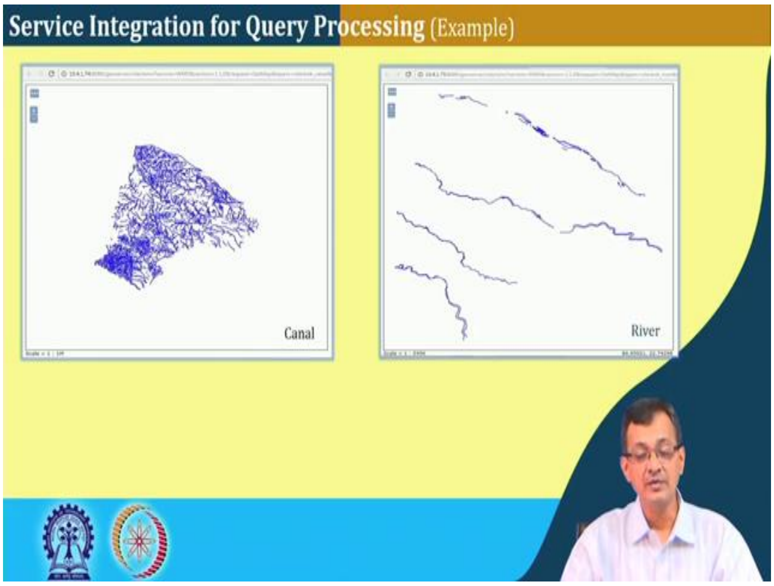
So, this is some sort of a graph type of things or we will sees at some times later in this course these are also considered as the spatial networks right. So, this is these are considered to be is a spatial network where it has different type of typical properties like you need to follow the road and type of things right.

So, I can say that a particular region if it is not connected or the connected to a round around way wait is far away. So, the Euclidean distance may be much near right you cannot scale a wall or type of things right you cannot just jump over the river, but you have to follow the path into the things right. So, that has a different type of properties we will see look into those things in some lectures right.

So, this is a this is my operation. So, I my operation was something shortest path related operation for that I require these are the data sets, if it is only road network then different sources of network if it is a multi modal type of transportation mechanisms then I have different data sets from road rail water body air etcetera and then those networks these transportation networks should come into a common schema so, that allow me to query this paths ok. So, that that is the thing for that I require this individual repositories need to be model. Model in a standard way create a standard way of machine readable things right then that XMI to XSD and then create the database at the underlining create the database and then whenever I require a particular things I populate this database and work on those query on those database.

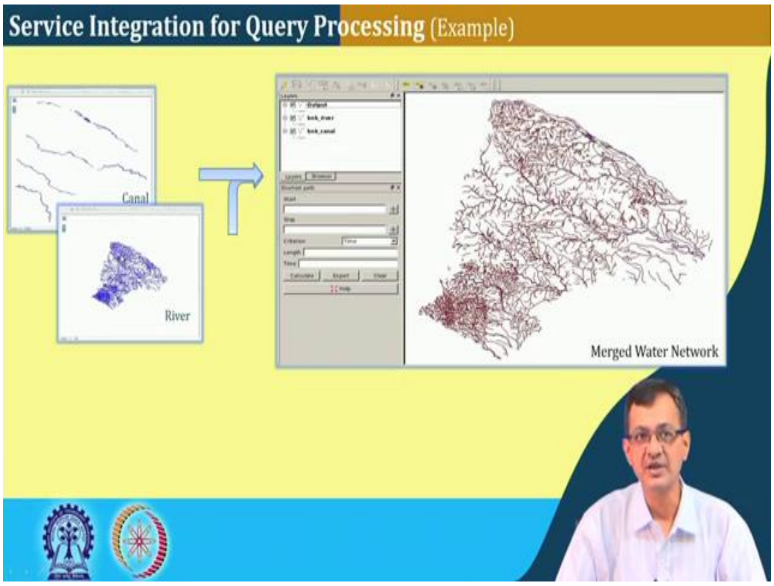
Because querying on a database is much faster and optimized right otherwise you have to do a long query which will be time consuming and may not results in a quick way.

(Refer Slide Time: 25:00)



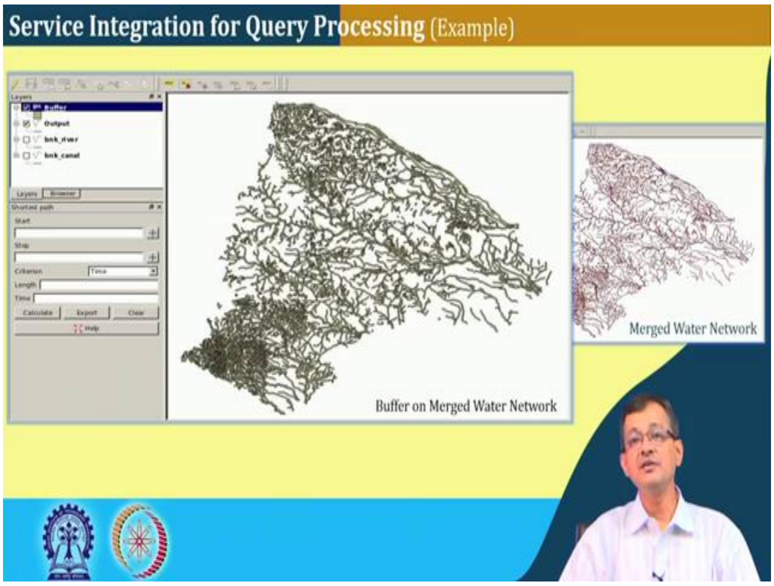
So, this is another that like canal network. So, another scenario. So, we have a river network and different type of canal network right. So, a particular region. So, there are different canals and then there are different type of river networks.

(Refer Slide Time: 25:24)



Now, I can have this sort of a thing again a merge water network. So, I have this sorry there is a mismatch a this problem in the labelling this should be river and this should be canal. So, this just swap this river and canal.

(Refer Slide Time: 25:43)

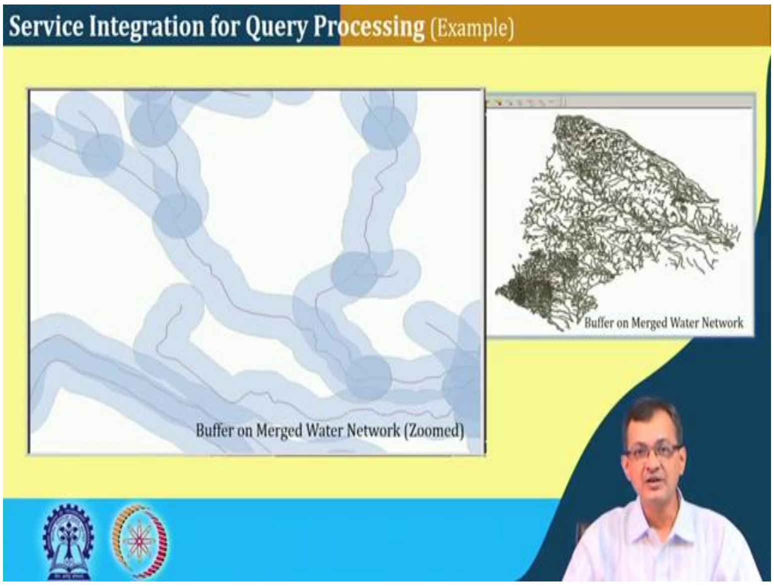


Now, we can have a something like a if you remember that some queries we are talking about, that if there is a flood situation then which are the areas likely to be under water right what are the areas likely to be inundated? Now what I require? What was my definition what was our definition of flood? So, the flood we define as that 1 kilometer or plus minus from a river network right. So, that is the way I define the flood. So, 1 kilometer plus minus and the river network.

Now, if it is that the definition of the flood then I need to have a river network, then create a buffer around it right a buffer means I a buffering on that river network or river or this river and canal water network we say and then find out that which are the reasons are affecting like I want to find out how much crop will be affected. So, the which are over that I can have land use land cover or vegetation map over that vegetation map these and then find out that how much things may be affected or which are the villages likely to. So, if the village boundaries are there, then I can overlay on the things and find out that which are the villages will be different so, that require a buffering operations.

So, couple of things first of all there is a river network, there is a canal network in our case these need to be integrated then create a buffer and then with that buffer I may need to integrate with the land use land cover or vegetation or village maps or block maps and type of things and find out the things. Now if this is my underlining queries the data structure or the data schemas should be in place so, that I can query on the things right here that is the that exactly we are trying to do.

(Refer Slide Time: 27:47)



So, little expanded form this is the buffer zoomed right. So, these are our network and there is plus minus something is there. There are actually things are not much more complicated because it is not like that only 1 kilometer or 500 meter plus minus it basically driven by the elevation model of the things as we are not considering immediately the 3D aspect. So, elevation model or even without the 3D elevation model if the elevation is there the water will flow in that direction right and that will have a different inundation pattern etcetera.

So, I require maybe require a hydrologic x part to find out a functional scheme that were how much inundation will be there and over that I put on if I want to find out that how much population will be affected then a population map over that this buffer or which are the villages likely to be affected over that things and so, and so, forth right. Or and directly somebody may give this inundation layer like from the from our some ISRO satellite, I can if that is given that this is my inundation layer right I can overlay and see that which other roads are still on which are the roads are off and type of things right.

So, this is this is this gives me a handling right. So, just to look into the things like. So, what we want at the end of the day is to query my things like as we started in well discussing or in somewhere in between, that we have some specific queries to answer right or information retrieval or data retrieval process has to be there in order to have this query to be answered I need the data right.

The data may be in different repositories which need to have a common denomination or common platform or I need a schema or structure where these queries can be satisfied for that we require these modeling is our quote unquote modeling is the first thing to start with were we have a model or standard modeling and then go on to create this schema. Then I can have the phase the data as and when required by the things.

Now once that is in a underlining database then querying on these databases becomes much easier or having different type of construct to do that. That we will see subsequently that how queries will be there and we will also see that this when we talk about this type of databases or spatial databases or spatial query, we know SQL, we know DBMS, that spatial DBMS spatial SQI. So, there are different challenges first of all this data load is pretty high right cannot be put on the main memory most of the cases.

So, how the query will work right. If you do not have this whole scenario of the whole graph on the memory then finding a shortest path will be a difficult problem right. So, how this query is a behave there is a geometry where it comes in to play picture if you when you talked about who discussed about different topological relationship, this geometry etcetera are important things right how these geometries will be addressed right.

So, this underlining database should understand this sort of spatial properties or spatial functionalities that is why we have spatial extension right like oracle spatial, MySQL spatial extension, POSIX or Postgres and there are different other type of things right that will subsequently discuss. And also that is important to get the data in the service bone that also will discuss so, that the service oriented architecture how they help me in doing so, fine. So, with this let us stop our discussion today and we will continue in the next class.

Thank you.